

## CLIPP

### Christiani Lehmanni inedita, publicanda, publicata

#### titulus

Information structure and grammaticalization

#### huius textus situs retis mundialis

[http://www.uni-erfurt.de/  
sprachwissenschaft/personal/lehmann/CL\\_Publ/  
information\\_structure\\_and\\_grammaticalization.pdf](http://www.uni-erfurt.de/sprachwissenschaft/personal/lehmann/CL_Publ/information_structure_and_grammaticalization.pdf)

#### dies manuscripti postremum modificati

16.07.2012

#### occasio orationis habitae

Conference on “New reflections on grammaticalization 3”,  
July 17 – 20, 2005, Santiago de Compostela

#### volumen publicationem continens

López-Couso, María José & Seoane Posse, Elena (eds.),  
*Theoretical and empirical issues in grammaticalization 3*.  
Amsterdam & Philadelphia: J. Benjamins (Typological  
Studies in Language, 77)

#### annus publicationis

2008

#### paginae

207-229

# Information structure and grammaticalization

Christian Lehmann

University of Erfurt

## Abstract

The contribution investigates how decisions made at the level of information structure predetermine the course of grammaticalization at various syntactic levels between the complex sentence and the clause, and how, on the other hand, grammaticalization levels out contrasts of information structure. Examples are taken, among others, from cleft sentences and relative constructions in Ancient Greek, Latin, French and Yucatec Maya.

## 1 Introduction

The slogan ‘from discourse to syntax’ (Sankoff & Brown 1976, Givón 1979) has been with us since the early days of work on grammaticalization.<sup>1</sup> The genesis of grammar – firstly, syntax – out of pre-grammatical, textual structure has become a cornerstone of the theory of grammaticalization. Among the topics that have received much attention in this connection are the grammaticalization of topic to subject (Shibatani 1991) and the condensation of combinations of sentences into one (Hackstein 2004). In this contribution, the focus will be on the information structure (alias functional sentence perspective) that obtains at the higher levels of grammatical structure. The issue is twofold: On the one hand, information structure is present in the source constructions that undergo grammaticalization and may direct their course. On the other hand, information structure is itself susceptible of grammaticalization. The following theses will be supported by empirical evidence:

1. Information structure is part of grammatical structure. Consequently, it may come about by grammaticalization, it may show different degrees of grammaticalization, and it may dissolve by grammaticalization.
2. The grammaticalization of any other element or construction may be conditioned by the information structure assigned to it.

## 2 Basic notions of information structure

### 2.1 The nature of information structure

The information structure of a sentence<sup>2</sup> concerns the way that the content of propositions is packaged in discourse depending on the speaker’s assessment of the current state of the universe of discourse. Information structure is part of linguistic structure, i.e. it has structural and functional aspects. As a **functional domain**, information structure comprises the following subdomains:

---

<sup>1</sup> I thank Knud Lambrecht for helpful criticism of a draft version.

<sup>2</sup> The following conception essentially stems from Lambrecht 1994.

1. presupposition vs. assertion of propositions (known vs. not yet known),
2. identifiability and activation of referents (status of mental representations in the addressee's mind, from active to brand-new unanchored),
3. topic vs. focus status of elements of propositions (predictability vs. unpredictability of relations between propositions and their elements).

The **structural aspects** are twofold. First, the propositions, predications and referents which are related in information structure are coded by lexical and grammatical means, which means that any knowledge that speakers and hearers may entertain which is not thus coded in a sentence is irrelevant for its information structure. Second, some of the concepts of information structure may be grammaticalized in individual languages, which means that there may be grammatical markers that correspond more or less closely to certain functions of information structure.

We will here chiefly be concerned with the topic – focus articulation of a sentence. The other two subdomains – presuppositions and the status of referents in the awareness of interlocutors – will be mentioned in passing.

A referent has the function of the **topic** of a proposition if, in that particular discourse, the proposition is about that referent (Lambrecht 1994:127, 131). A proposition may be about more than one referent; thus, a sentence (such as E1 below) may have more than one topic. A topic must be in the universe of discourse; a topic expression therefore must be a referring expression. There is a **Topic Accessibility Scale** according to which a referent is the better suited as a topic, the more active it is in the awareness of the interlocutors.

The typical topic expression is a clitic personal pronoun. Its position in the sentence is determined by its syntactic and semantic function; it does not matter for information structure.<sup>3</sup> If a referent is not sufficiently active, it has to be evoked by a lexical NP. However, for it to be the topic of its clause would violate the **Principle of the Separation of Reference and Role** (PSRR, Lambrecht 1994:185): “Do not introduce a referent and talk about it in the same clause.” In such cases, the topic is split into a left-dislocated NP which announces the topic and a resumptive pronoun that represents the topic for the predication. The lexical NP typically occupies the position preceding the initial sentence boundary. Constructions such as left-dislocation serve the purpose of making referents available as topics that would not, at the current moment in the discourse, be sufficiently high on the Topic Accessibility Scale and which therefore need special introduction.<sup>4</sup> This syntactic separation obeys the PSRR. E1 illustrates two things at once: First, a resumed topic which consists of the left-dislocated *i bambini* and the resumptive pronoun *li*; second, the two topics *i bambini* and *quando piove*.

E1. I bambini quando piove li portiamo a scuola in macchina. (Banti 2005:2)  
 ITAL ‘The kids, when it rains, we take them to school by car.’

The meaning of a sentence is represented by a set of propositions that may be subdivided into presupposition and assertion. Given this, the **focus** may be defined (Lambrecht 1994:213) as “the semantic component of a pragmatically structured proposition whereby the assertion ... differs from

---

<sup>3</sup> Lambrecht (1994:202) argues that the position of pronominal expressions in the sentence tends to be fixed by rules of grammar and is therefore not amenable to functional explanations in terms of information structure.

<sup>4</sup> Left-dislocation serves “to promote a referent on the Topic Accessibility Scale from accessible to active status, from which point on it can be coded as a preferred topic expression, i.e. as an unaccented pronominal.” (Lambrecht 1994:183)

the presupposition.”<sup>5</sup> The focus is, thus, the central part of the assertion of a sentence, just as the topic is the central part of the presupposition. This is visualized in S1.

S1. *Central notions of information structure (Lambrecht 1994)*

proposition	presupposition	assertion
core component	topic	focus

Both topic and focus may be contrastive. A construction with a contrastive topic X may be paraphrased by *as for X, ...*, while a construction with a contrastive focus X may be paraphrased by *it is X that ...*

E2. Who married whom? – John married Linda, and Paul married Sue.

In E2, *John* and *Paul* are contrastive topics, *Linda* and *Sue* are contrastive foci. This can be proven by the paraphrase test:

E2'. Who married whom? – As for John, it was Linda he married, while as regards Paul, it was Sue he married.

At the same time, constructions such as these, which transcend the clause level, are often the input to the grammaticalization of topic and focus.

## 2.2 Relations between information structure and grammaticalization

The two main sections of this paper will be devoted to the interaction of information structure with grammaticalization. These two concepts belong to distinct theoretical levels, and their interaction is therefore no straightforward matter. Two entirely different kinds of interdependence of the two domains may be deduced from their nature.

As seen in section 2.1, information structure is a notion that has both functional/semantic and structural aspects, just as grammatical categories like tense and syntactic functions like the indirect object. These categories and functions are subject to grammaticalization, i.e. they can evolve and dissolve into something different through grammaticalization. Their genesis is semantically motivated; but the more they are grammaticalized, the more they become arbitrary components of language-specific grammatical structure. All of that applies, in principle, to the notions of information structure, too. The three dimensions enumerated in section 2.1 are primarily universal communicative functions which may or may not be fulfilled in a language by particular grammatical structures; and the latter may evolve and change by grammaticalization. The main difference between the notions underlying those grammatical categories and functions, on the one hand, and the notions underlying information structure, on the other, is that the former exclusively or at least partially serve the cognitive function of language, while the latter only serve the communicative function of language. Consequently, grammaticalization of these two main functional domains draws on different kinds of input: Items and constructions serving the cognitive function of language are mainly recruited among the lexical items of the language, while the temporal order of chunks of utterances reflecting the order of communication, and prosody reflecting the relief of ongoing thought and argument, are much more prominent in the genesis of constructions that serve the communicative function of language. Although formatives such as focus markers and topic markers do evolve by grammaticalization of lexical items, the evolution of sheer formal structure – syntactic and morphological constructions with their sequential and prosodic properties – out of information structure is

---

<sup>5</sup> It probably follows from this definition that the focus is that part of a sentence that is in the scope of such highest-level operators as illocutionary force, negation and quantification (Banti 2005:3).

much more important in this area. The grammaticalization of information structure will be the topic of section 3.

The second way that information structure interacts with grammaticalization follows from the asymmetry constitutive of the three dimensions enumerated in section 2.1. Each of these dimensions is bipolar, with one member being prominent in speech and demanding the hearer's attention, the other member providing the background. The first member is elaborated in linguistic expression, the second member may be reduced. In grammaticalization, then, the communicative asymmetry is reflected in syntactic dependence where pragmatically backgrounded material attaches to pragmatically foregrounded material. The conditioning of grammaticalization by information structure will be the topic of section 4.

### 3 Grammaticalization of information structure

#### 3.1 Contrastive-focus cleft sentence

##### 3.1.1 Incipient grammaticalization

The operations of information structure are highly abstract. They deal with the manipulation of the universe of discourse and of the flow of information into the mind of the hearer, not with concepts that may denote anything.<sup>6</sup> The grammatical means employed in marking information structure are therefore more subtle from the beginning. They are seldom recruited from among lexical items. In E3, a new topic is announced explicitly by a circumlocution involving the lexical verb *concern*. Similarly in E4, the extrafocal clause of a suspension focus is introduced by a lexical verb such as *know* or *guess*.

E3. As far as the king of France is concerned, he is bald.

E4.  $\left\{ \begin{array}{l} \text{You know} \\ \text{Guess} \end{array} \right\}$  who is coming for dinner tonight: the king of France!

Such verbs make explicit the relation of the referent to the predication or to the hearer's consciousness. The constructions of E3 and E4 are not (yet) specific grammatical constructions dedicated to a particular information structure. They do, however, share with the latter the immaterial features mentioned, viz. order of components and prosody. In the following subsections, we will see the genesis of two kinds of cleft sentence out of regular syntactic constructions.

The most explicit syntactic strategy of contrastive focus is sentence-clefting. It involves the formation of a complex sentence of the structure of S<sub>3</sub> in S<sub>2</sub>.

S<sub>2</sub>. *Cleft sentence*

[ [ Δ	X	]s <sub>1</sub>	[ ...	]s <sub>2</sub>	]s <sub>3</sub>
expletive/zero subject	non-verbal predicate				
	focus expression		extrafocal clause		
[ non-verbal clause		]	[ open clause		]
[ main clause		]	[ dependent clause		]

The subject of the main clause is semantically empty. It depends on the syntax of the language whether it requires an expletive subject there, just as it depends on the language whether it requires a

<sup>6</sup> The relation of denotation is between a sign, including its concept, and a physical object. For instance, *apple* has a denotation, *presupposition* has no denotation.

copula with non-verbal predicates.<sup>7</sup> If both are missing, the main clause of a cleft sentence reduces to the focused constituent. The extrafocal clause commonly takes the form of a complement clause subordinated by the universal subordinator ('that'). It is open in the sense that one of its arguments or satellites, viz. X, is missing. The order of S<sub>1</sub> and S<sub>2</sub> is not crucial, although it is usually that order.<sup>8</sup>

The cleft sentence is the grammatical construction functioning as argument focus construction in many languages, among them French, where it is practically the only construction available for that information structure, and Italian, where it is optional. The construction is partly motivated: It reflects the separation of focus and presupposition by a two-clause structure, and it reflects the attention cline between focus and presupposition by the asymmetric syntactic status of main clause and dependent clause. Finally, it puts the focus expression, and nothing else, into the predicate of the main clause, thus assigning the focus the canonical syntactic function for new information. However, not all aspects of the construction are motivated: Neither of the two clauses composing it has itself argument focus. That is, instead, a non-compositional semantic property of the grammatical construction 'cleft sentence' as a whole (Lambrecht 1994:230). To that extent, the cleft-construction evinces incipient grammaticalization.

The notions articulating information structure which appear in S1 are not autonomous components that could be identified with sentence constituents. Information structure is relational (Lambrecht 1994): all of it is relative to a certain speech situation; the assertion is relative to its presupposition; something is more or less topical; and the focus is the difference between presupposition and assertion. Two guiding principles follow from this:

1) It is not possible for a particular component of information structure to be grammaticalized in isolation. What can be grammaticalized is a certain articulation of information structure.

2) Grammaticalization of information structure means that pragmatic relations lose their specificity, that differences between pragmatic components are leveled out.

Thus, the topic cannot be grammaticalized; but the cline between topic and comment can be leveled out. Similarly, the difference between focus and presupposition can be smoothed out.

In French, the focus cleft sentence displays the structure of S2. Although it is constructed according to high-level and general syntactic rules, it already exhibits some symptoms of grammaticalization.

- E5. a. C'est qui? – Ce sont les étudiants.  
FR 'Who is that? – Those are / It's the students.'
- b. C'est les étudiants qui ont raison.  
'It's the students who are right.'

In the simple copula sentence E5.a, the verb agrees in number with its complement. In the main clause of a cleft sentence (E5.b), it does not. This points to a loss of grammatical differentiation, thus to some degree of grammaticalization of the cleft sentence.

### 3.1.2 The leveling of focality

The 'rule of accommodation for presupposition' (Lambrecht 1994:67) says: If an expression requiring a certain presupposition is actually used, then it forces that presupposition, i.e. after the utterance

<sup>7</sup> The term 'non-verbal clause' is intended to cover clauses both with a nominal and with a copula predicate.

<sup>8</sup> This means that the distinction between clefting and pseudo-clefting is not made on the basis of clause order. Instead, the criterion is the nature of the extrafocal clause: in the case of pseudo-clefting, it is a free relative clause.

the presupposed proposition is in the universe of discourse, whether the hearer had entertained that proposition previously or not. In other words, if the hearer did not entertain that proposition before, he is asked by the utterance to accommodate it. On the basis of the rule of accommodation, a presupposition may be conventionalized and grammaticalized.<sup>9</sup>

A cleft sentence of the general structure [ [  $\Delta$  is X ] [ extrafocal clause ] ] (cf. S2) presupposes that the extrafocal clause applies to some Y and asserts that Y = X. By the rule of accommodation, a cleft sentence may be used in a situation where the presupposition of the extrafocal clause is not shared. The audience is then asked to accommodate that presupposition. If that use of the construction becomes conventionalized, it becomes “an indirect way of communicating the content of that proposition” (Lambrecht 1994:71). In a sentence such as E6,<sup>10</sup>

E6. c'est Jean-Paul Sartre qui a dit que l'homme était condamné à inventer l'homme.  
FR (www.beurfm.net/forums/viewtopic.php, Juin 8, 2004)  
'It's Jean-Paul Sartre who said that man was doomed to invent man.'

nothing in the surrounding context suggests that anybody said the proposition of the extrafocal clause. To the extent that pragmatic accommodation of the proposition presupposed by the extrafocal clause is conventionalized in the cleft sentence, this construction becomes more grammaticalized<sup>11</sup> and ceases to always code contrastive focus.

### 3.2 Presentational cleft sentence

The French presentational cleft sentence codes an all-new utterance, as in E7.

E7. Y a le téléphone qui sonne!  
FR 'The telephone is ringing!' (Lambrecht 1994:194)

In such an utterance, not only the predication on the referent, but the referent itself is new. The utterance therefore has to fulfil two functions at the same time, which should be separated according to the PSRR. A rather explicit and iconic way of doing this is to first introduce the referent and then to make the relevant predication. This is just what underlies the French *y a*-cleft construction of E7. It consists of two clauses: first, an existence predication introduces the referent; second, a relative clause provides the predication. The subordination of the predication under the introduction of the referent calls attention to the fact that the latter is not a topic for the predication. Now this construction is grammaticalized in colloquial French (cf. Lambrecht 1994:234). Symptoms of this include the following:

- There is apparently an existence predication which, however, does not assert the existence of anything;
- the complement of the existence predication may be definite, as in E7;
- by structure and intonation, the relative clause looks like a restrictive one, but it could never be semantically restrictive, given the definiteness of its antecedent.<sup>12</sup>

<sup>9</sup> The following account stems from Lambrecht 1994:70f.

<sup>10</sup> Such examples are legion; cf. *C'est Voltaire qui a dit: "Un lion mort ne vaut pas un moucheron qui respire"*. (www.simimpact.com/, 10.07.05).

<sup>11</sup> Commenting on similar examples from Italian, Berretta 2002, §5.2 finds that the cleft-sentence is losing its focusing force. Hyman (1984:80) generalizes that “languages, through their grammars, could ‘harness’ the pragmatics and create a formal system for focus”.

<sup>12</sup> The antecedent of a restrictive relative clause is indeterminate (neither definite nor generic). In a restrictive relative construction such as *le téléphone qui sonne* ‘the telephone that is ringing’, the antecedent of the

To that extent, the French presentational cleft construction has been desemanticized. The two clauses are not construed separately and compositionally, and instead the first clause only contributes a referent, and the second clause only contributes a predication, to the unitary act of making athetic statement. Again, like the focus cleft sentence treated before, the presentational cleft sentence evinces incipient grammaticalization.

### 3.3 Reduction of emphasis

Emphasis arises when there is a conflict between presupposition and assertion. One function of periphrastic *do* in English is to mark emphasis in this sense. The following account stems from Lambrecht 1994:71f. Emphatic *do* in English is grammaticalized in three stages:

1. Fully contrastive contradiction:

E8. I did pay you back.

E8 involves a presupposition ‘it is doubted that speaker paid addressee back’, which is contradicted by the assertion.

2. Attenuated contradiction:

E9. I was afraid to hit him; I did insult him, though.

E9 involves a presupposition ‘one might think that I did not even insult him’. Again, the assertion does contradict it, but as the presupposition itself is weaker, the contradiction is, too.

3. Non-contradictory emphasis:

E10. I do hope that doggie’s for sale (song line)

E10 involves no presupposition suggesting that the speaker does not hope that the dog is for sale. There is just a weak emphasis on the verb, paraphrasable by *I really hope*. Lambrecht (1994:72) says “the emphatic *do*-construction has become a conventionalized grammatical way of expressing emphasis.” The semantic aspect of the grammaticalization through stages 1 – 3 is the gradual loss of a presupposition. The history of emphatic *do* is, thus, a story of the leveling of information structure.

## 4 Information structure conditioning grammaticalization

### 4.1 Relative clauses

A textual combination of two clauses can coalesce into a complex sentence if they are related at the level of semantics and of information structure.<sup>13</sup> Two kinds of semantic connection between the two clauses are of relevance for their coalescence, interpropositional and anaphoric (technically, endophoric) relations. Here we will concentrate on the latter. If the connection between two adjacent clauses is by anaphora, then information structure may shape the functioning of anaphora, and the pronouns will get different functions depending on whether they prepare a topic to be resumed (as *kwit* in E12 below) or they resume an existent topic (as *tòn* in E16).

At the start, the two clauses are at the same syntactic level, in that sense the construction is symmetric. Combination of the two clauses into a sentence leads to parataxis. However, such struc-

---

relative clause is *téléphone*, not *le téléphone*. (cf. Lehmann 1984, ch. V.2). In E7, however, the antecedent of *qui* is actually *le téléphone*.

<sup>13</sup> This idea is traced back to the 19<sup>th</sup> century in Hackstein 2004, section 1.

tures have the potential of becoming asymmetric by grammaticalization. Then the question arises which of the two clauses becomes subordinate to the other. As we will see, this alternative is essentially decided by the information structure assigned to such complexes.

We start from a textual combination of two clauses  $S_1$  and  $S_2$  that are in an anaphoric relation such that the first introduces a referent  $X_i$  that the second resumes by  $Y_i$ . The general constellation is depicted in S3.<sup>14</sup>

S3. *Anaphoric connection*

[ ...  $X_i$  ... ]<sub>S1</sub> [  $Y_i$  ... ]<sub>S2</sub>

As an example, consider E11, which is an ambivalent translation of the Hittite example E12 below:

E11. From the campaign I brought some booty; with that I adorned them.

- a. (As for) the booty I brought home from the campaign, with that I adorned them.
- b. From the campaign I brought some booty, with which I adorned them.

The textual sequence of the two clauses of E11 may be used, in a particular discourse, with the information structure of version (a) or of version (b).

#### 4.1.1 The preposed relative clause

Now there are two variants of the constellation S3 which differ in their information structure. In the first alternative,  $X_i$  is indefinite and specific and is introduced as something to be resumed. The whole clause  $S_1$  serves to characterize and possibly even identify this referent.  $S_1$  is a left-dislocated topic. Proto-Indo-European marks this information status of an NP by the clitic indefinite pronoun \**kwis*.<sup>15</sup> Since  $S_2$  takes  $X_i$  thus characterized up anaphorically by  $Y_i$ ,  $S_1$  as a whole constitutes an external topic for  $S_2$ . This interpretation of E11 is made explicit in E11.a. A natural text example is E12 from Hittite.

E12. KASKALz-a kwit assu utahhun  
HITT campaign:ABL IND:ACC.SG.INAN booty bring.home:PST:1.SG

n-at apedanda halissiyānūn.  
CONN-3.INAN.ACC D3-INST adorn:PST.1.SG

‘With the booty that I had brought home from the campaign, I adorned them.’ (ap. Lehmann 1984:179)

The construction is called the **correlative diptych**. Now this is the dominant variant of the Old Hittite relative construction. The pronoun *kwis*, marking the  $X_i$  of S3, is already a relative pronoun. The overwhelming majority of the relative clauses of the text corpus are preposed like E12; a few are postposed or even postnominal. The latter variants will not be illustrated from Hittite; the Hittite data merely serve to show that the preposed relative clause belongs to a stage that was common to Hittite and Latin. E13 is an example from Old Latin.

E13. ab arbore abs terra pulli qui nascentur,  
LAT from tree:ABL.SG from ground:ABL.SG off-shoot:NOM.PL IND:NOM.PL be.born:FUT:3.PL

<sup>14</sup> The idea of alternate information structures giving rise to different relative constructions is put forward in Sankoff & Brown 1976. Cf. also LaPolla 1995:316-318 on the information structure difference between a prenominal relative clause and a postnominal “descriptive clause” in Mandarin Chinese. The analyses presented below are from Lehmann 1984, ch. VI.1.1.

<sup>15</sup> Lambrecht 1994:83 mentions a couple of languages which use a numeral classifier to mark this function.

eos in terram deprimito  
 them in ground:ACC.SG push.down:IMP  
 (originally:) ‘from the tree there will be off-shoots growing up from the ground; those you have to push back into the ground’; (synchronically:) ‘the off-shoots of the tree that will grow up from the ground have to be pushed back into the ground’ (Cato *agr.* 51)

If the correlative diptych were the only type of relative construction to be found in the Latin text corpus, we would probably not feel entitled to call the indefinite pronoun a relative pronoun.<sup>16</sup> However, this particular relative construction is gradually ousted by its positional variants; it does not survive into a single Romance language. The first step in its grammaticalization is the interpretation of the preposed clause as a clausal NP. This reanalysis can be verified if that NP is semantically definite in the discourse context, as it must be in E14.

E14. Quae mihi antea signa misisti,  
 LAT REL:ACC.PL.N me:DAT before statue:ACC.PL send:PRF:2.SG  
 ea nondum vidi.  
 it:ACC.PL.N not:yet see:PRF:1.SG  
 ‘The statues you sent me the other day, I have not seen yet.’ (Cic. Att. 1, 4, 3)

E14 contains another symptom of the subordination of the first clause under the second, viz. the initial position of the relative pronoun. The clitic indefinite pronoun could not yet introduce a clause; but the relative pronoun now functions, at the same time, as a subordinator.

In E14, the relative clause is still left-dislocated and serves as the topic for the subsequent main clause. This is not so in E15.

E15. Cave tu idem faxis  
 LAT beware you.NOM.SG ACC.SG.N:same do:PRF.SUBJ:2.SG  
 alii quod servi solent!  
 other:NOM.PL REL:ACC.SG.N slave:NOM.PL use:3.PL  
 ‘Don’t you do the same which the other slaves tend to do!’ (Pl. As. 256)

In E15, the relative clause codes an accommodated presupposition (cf. section 3.1.2), is thus thematically backgrounded, and it is postposed. The construction is called the **inverted diptych**. It is the diachronic basis of the postnominal relative clause, which we may illustrate by a somewhat clumsy, non-Ciceronian variant on E14:

E14’. Nondum signa quae mihi antea misisti vidi.  
 ‘I have not yet seen the statues you sent me.’ (CL)

In E15 and E14’, the information structure that motivated the formation of the relative construction in the first place is leveled in the sense that the relative clause may play any role in the information structure of its matrix sentence. Thus, this is an example of the genesis of a particular syntactic construction on the basis of a textual combination of two clauses pre-structured by information structure which is subsequently leveled out.

#### 4.1.2 The postposed relative clause

In the other assignment of information structure to E11, summarized in E11.b, S<sub>1</sub> of S<sub>3</sub> contains a comment on the referent of X<sub>i</sub> that may already be in the universe of discourse, while S<sub>2</sub> makes an

<sup>16</sup> Given, however, the fact that the indefinite pronoun had been reinforced (*ali-qui*) in Old Latin, the original interpretation of E13 must be regarded as reconstructed.

additional comment on that referent, resuming it with the definite pronoun  $Y_i$ . Proto-Indo-European had a weakly demonstrative personal pronoun *\*so-/to-* ‘that’, which was typically used in the function of  $Y_i$  of S3. It is illustrated in E16 from Homeric Greek.

- E16. Allà tí moi tôn êdos,  
 AGR but INT.NOM.SG.N I:DAT DEM:GEN.PL pleasure:NOM.SG.N
- epeì phílos ôleth’ hetaîros,  
 since dear:NOM.SG.M perish:PST.3.SG friend:NOM.SG.M
- Pátroklos, tôn egō perì pántōn  
 Patroklos:NOM.SG.M [DEM:ACC.SG.M I over all:GEN.PL
- tîon hetaírōn îson emêi kephalêi;  
 esteem:PST.1.SG friend:GEN.PL.M like:ACC.SG.M my:DAT.SG.F head:DAT.SG.F ]
- tôn apôlesa.  
 DEM:ACC.SG.M loose:AOR:1.SG
- ‘But what pleasure do I yet have in those things after my dear friend perished, Patroklos, whom I esteemed over all friends like my head; him I lost.’ (Hom. *Il.* 18, 80-82)

By their structure, both of the clauses introduced by *tôn* ‘him, that one’ may be independent. The first (though not the second) of them may as well be construed as a relative clause. It would be non-restrictive since it could not form a narrower concept of its head noun. This construction is, thus, on the threshold between paratactic anaphoric clause and relative clause.

In E17, the second clause can still be interpreted as independent. Here, however, the antecedent is indeterminate. Thus, the second clause may form a more specific concept on the basis of the antecedent concept (as there are islands not surrounded by the ocean); so if taken as a subordinate clause, it could be a restrictive relative clause.

- E17. eîdon gàr nêson  
 AGR see:PST.1.SG for island:ACC.SG.F
- tên péri póntos apeíratos estephánōtai  
 [DEM:ACC.SG.F around sea:NOM.SG.M infinite:NOM.SG.M surround:3.SG]
- ‘for I saw an island which is surrounded by the infinite ocean’ (Hom. *Od.* 10, 194f)

In E18, the head noun *éridos* together with the relative clause identifies a referent that is established in the universe of discourse. Here, the second clause can no longer alternatively be construed as an independent comment on that referent.

- E18. oud’ Agamémnōn lêg’ éridos  
 AGR NEG:however Agamemnon:NOM.SG.M desist:PST.3.SG quarrel:GEN.SG.F
- tên prōton epēpeílēs’ Akhillēi  
 [DEM:ACC.SG.F first threaten:AOR.3.SG Achilles:DAT.SG.M]
- ‘Agamemnon, however, did not let go the quarrel that he had earlier threatened Achilles with’ (Hom. *Il.* 1, 318f)

We now have a restrictive relative clause introduced by the relative pronoun *ho-/to-*. It is no longer restricted to additional comments on a pre-established referent. In E19, the NP containing such a relative clause is even a left-dislocated topic.

- E19. tās dè stēlās tās histā ... Sésōstris,  
 AGR DET:ACC.PL.F however column:ACC.PL.F [REL:ACC.PL.F erect:PST.3.SG Sesostris:NOM.SG.M]
- hai mèn pleûnes oukéti phaínontai perieoùsai

DET:NOM.PL.F Ø most:NOM.PL.F not:still appear:3.PL be.left:PTCP.PRS:NOM.PL.F

‘most of the columns that Sesostris had put up do no longer seem to be extant’ (Hdt. 2, 106)

Observe that at the end of this development, what started out as a postposed relative clause ends up in the same kind of construction which is at the beginning of the evolution of the preposed relative clause (cf. E12). We found just the converse to be true for the original preposed relative clause (E15). This shows that a syntactic construction may be tied up with a particular information structure at its genesis. In the course of its grammaticalization, however, the information structure inherent in the construction is leveled out, which means that the grammaticalized construction becomes compatible with diverse information structures which originate independently from it. The initial function of the construction and its elements in information structure does not persist until the end; there comes a point where the origin becomes unimportant, and constructions of opposite origin may fulfill the same grammatical function.

## 4.2 Pronominal interrogatives

We take up the topic of the cleft sentence already touched upon in section 3.1. Here, however, the issue is not the grammaticalization of the focus construction but instead the grammaticalization of the pronominal interrogative. The basic illocutionary force of this construction derives from marking an indefinite pronoun as focus, which thereby becomes an interrogative pronoun. In many languages, sentence-initial position of the interrogative pronoun suffices to get that effect, as in E20 from Classical Latin.

E20. a. quis nescit ...

LAT ‘who does not know ...?’ (Cic. *de orat.* 2, 62, 5)

b. quid fieri potest ...?

‘what can be done ...?’ (Iust. *Dig.* 34, 2, 6, 1, 5)

In Latin, the cleft sentence is not a grammatical construction. Sentences that look like it are regular relative constructions. However, such constructions do occur with some regularity if what is in focus is an interrogative pronoun, as in E21.

E21. a. quis est qui nesciat ...

LAT ‘who does not know ...?’ (Cic. *de orat.* 2, 45, 4)

b. Quid est igitur quod fieri possit?

‘What then can be done?’ (Cic. *Verr.* 1, 1, 32)

Since the cleft sentence is not grammaticalized, such constructions are maximally emphatic.

On the way towards French, the interrogative pronoun itself loses in substance, and the cleft sentence gradually becomes a dedicated grammatical construction for interrogative pronoun questions. Here are two examples from Old French (apud Rouquier 2002:101, 110):

E22. Qui est ce, diex, qui m’aparole?

OFr ‘Who is it, oh gods, who is speaking to me?’ (*Renart IV*, 233)

E23. Et savez vos que ce est que vos m’avez otroié?

OFr ‘And do you know what it is that you have empowered me to do?’ (*Mort Artu* 14, 13)

E22 has a vocative between the focus and the extrafocal clause, while E23 has subordinate clause word order in the interrogative clause. This shows that the cleft interrogative was not yet fully grammaticalized at that stage. It still marks some kind of insistence (Rouquier 2002:100f).

E24. C'est qui me parle?  
FR 'Who is speaking to me?'

In Modern French as illustrated by E24, the cleft sentence is the default and often the only possible construction for most interrogative pronoun questions. While the erstwhile structural apparatus of the cleft-construction becomes part of a renovated interrogative pronoun ([kjɛski] 'who', [kɛskə] 'what'),<sup>17</sup> the particular focus on the questioned constituent vanishes. The relief of the information structure is flattened and reduced to the standard focus associated with pronominal interrogatives.

Before we conclude, we look back at the Latin stage. The collocation *quis est qui / quid est quod* that undergoes the grammaticalization illustrated by E21 – E24 regularly occurs in another context, too, viz. with the meaning 'there is anyone who / anything which', typically after the conditional conjunction, as in E25.

E25. a. si quis est qui his delectetur ...  
LAT 'if there is anybody delighted by such things ...' (Cic. Tusc. 5, 102, 3)  
b. si quid est quod fieri possit ...  
'if there is anything that can be done ...' (Cic. Att. 11, 16, 5, 5)

The resemblance between E21 and E25 is superficial: First, the pronoun *quis/quid* has interrogative force in E21, but is an indefinite pronoun in E25. Second, the verb *esse* is a copula in E21, but a verb of existence in E25. This could not be the other way around: the focus construction of E21 is associated with identification, while the indefinite quantification of E25 is associated with existence. In a pronominal interrogative, it is the identification of the indefinite which is at stake. This is highlighted by the cleft construction. It is this information structure which then provides the frame for the grammaticalization of the construction.

### 4.3 Focused progressive aspect

In principle, the X representing the focus in S<sub>2</sub> may be any constituent of S<sub>2</sub>. However, the finite verb of S<sub>2</sub> cannot easily be focused in a cleft-construction, for at least two reasons. First, the predicate of S<sub>1</sub> cannot be a finite verb because otherwise its minimal syntactic structure (which requires a nominal predicate) could not be guaranteed. Second, by clefting its place in S<sub>2</sub> would become empty, so that S<sub>2</sub> would no longer be a clause. The solution to the first problem is to nominalize the focused verb. The solution to the second problem is to represent it by a pro-verb in S<sub>2</sub>, which is commonly a verb meaning 'do'. E26 illustrates this strategy from English.

E26. lying is what he does in between eating babies and denying old people their social security.  
(www.phantasytour.com/phish/ boards\_thread.cgi, 12.07.05)

The primary function of **verb focus** is to concentrate on the meaning of the verb itself while relegating everything depending on it to the presupposition. Nominalizing it serves this purpose well because nominalization generally involves valence reduction.

The English verb focus construction results from a regular combination of independently existing syntactic operations, i.e. it is not grammaticalized as such. In Colonial Yucatec Maya, sentence-clefting takes the general form of S<sub>2</sub>, as illustrated by E27.

<sup>17</sup> According to Lambrecht (p.c.), speakers strongly tend to say *qu'est-ce qui* instead of *qui est-ce qui*, i.e. in this particular environment, the *qui - que* contrast is neutralized. Reduction of the paradigm is another symptom of grammaticalization.

E27. ma' lawa lawa a mènent-ik-e'x in kol !  
 CYM [ NEG at.random ]<sub>S1</sub> [ SBJ.2 make-DEP-2.PL POSS.1.SG milpa ]<sub>S2</sub>  
 'don't make my cornfield at haphazard!' (*Cordemex s.v. lawak bik*)

Main clause and extrafocal clause according to S2 are indexed in the interlinear gloss. Yucatec Maya is one of those languages that have neither a dummy subject nor a copula nor a universal subordinator. The only feature of E27 to prove that it is a complex sentence is the dependent status of the extrafocal verb.

The verb focus construction follows the same pattern. I will first illustrate it from Modern Yucatec Maya, because it involves the same verb as a pro-verb that happens to function as a full verb in E27.

E28. a. h kíim-ih  
 MYM PST die-CMPL(SBJ.3)  
 'he/she/it died'

b. kíim-il t-u mènent-ah  
 die-INCMPL PST-SBJ.3 make-CMPL  
 'what happened is he died (lit.: dying is what he did)' (HK'AN 620.3)

The focused verb goes into the incompletive status, which is morphologically identical to a nominalized verb. The extrafocal clause of E28.b has an aspect of its own. Its verb is the Modern Yucatec Maya verb meaning 'make, do'. The construction is fully productive.

The verb focus construction with this particular pro-verb was already available in Colonial Yucatec Maya. However, in addition the language possessed a construction that was syntactically similar, except that it used a different pro-verb. There was a verb *kib* 'do' which was totally irregular. Its dependent status is suppletive *ka'h*, as appears in E29, which is itself a focus construction (although not a verb-focus construction).

E29. balam-il u ka'h Pedro  
 CYM tiger-ADVR SBJ.3 do.DEP Peter  
 'Peter makes the tiger / Peter is like a tiger' (lit.: 'tiger-like is what Peter does'; Arzápalo Marín 1995, III s.v. *cah*<sub>3</sub>)

This is the verb chiefly used in Colonial Yucatec Maya verb-focus constructions, as in E30.

E30. kambes-ah in ka'h ti' pàal-alo'b  
 CYM teach-INTROV.INCMPL SBJ.1.SG do.DEP LOC child-PL  
 'I am teaching the children' (Coronel 1620:72)

The information structure of the last two examples is the same. The syntactic structure is essentially the same, too, except that the focus in E30 is the main verb that is missing from its extrafocal clause. The internal syntax of the extrafocal clause has been adapted in that what was the direct object of the verb has become a prepositional object. No such adaptations are necessary if the focused verb is intransitive. The verb focus construction is, thus, marked with plurivalent verbs.

As already suggested by the translation of E30, the same construction functions as the progressive in Colonial Yucatec Maya. The paradigm shown in E31 may be gathered from the first colonial grammar (Coronel 1620:71f).

E31. a. hàan-al bín u kib  
 CYM eat-INCMPL go SBJ.3 do(SUBJ)  
 'he is going to eat'

- b. lúub-ul            t-u            kib-ah  
 fall-INCMPL    PST-SBJ.3    do-CMPL  
 'he fell'
- c. hàan-al            u            ka'h  
 eat-INCMPL    SBJ.3 do.DEP  
 'he is eating'
- d. hàan-al            u            kib-ah  
 eat-INCMPL    SBJ.3 do-CMPL<sup>18</sup>  
 'he was eating'

It seems that when pure verb focus is intended, the extrafocal clause may show its own aspectual auxiliary, with corresponding verb statuses (E31.a and b), whereas the progressive reading results if no aspect is marked (E31.c and d).

The progressive aspect views what the verb designates as an ongoing situation that the referent of the subject is in. Consequently, the functional locus of the progressive aspect is in intransitive verbs.<sup>19</sup> The verb focus construction is therefore well suited to get grammaticalized into a progressive aspect. The resulting construction may be dubbed **focused progressive**.

In Modern Yucatec Maya, the progressive category has been renewed on a different structural basis that we can forego here. However, it will be interesting to pursue the fate of the focused progressive construction. The pro-verb *kib* is fossilized in Modern Yucatec; only the form *ka'h* occurs in a couple of contexts. The general verb meaning 'make, do' is now *mèent* (with its variant *bèet*); and it is employed in the verb focus construction (as in E28.b). The Colonial focused progressive construction only survives in the modern immediate future, illustrated by E32 (to be compared with E31.a).

- E32. bîin    in            ka'h    kíim-il  
 MYM go    SBJ.1.SG    do.DEP    die-INCMPL  
 'I am going to die' (FCP 395)

The interlinear gloss of E32 is etymological. It makes explicit a focused progressive where the verb *bîin* 'go' is focused. What remains in the extrafocal clause would depend on the focused verb if that stayed in its clause. The grammaticalization of the construction involves, among others, the following steps:

- *bîin* 'go' is semantically bleached.
- The incomplete verb remaining in the extrafocal clause is reinterpreted as the main verb.
- The internal structure of the complex '*bîin* subject\_clitic *ka'h*' is blurred; it is reanalyzed as a discontinuous immediate future auxiliary with internal inflection.
- The whole sentence ceases to be complex; it is reinterpreted as a single clause.
- Whatever may have remained of the focal emphasis on the initial verb vanishes; the construction becomes open to different information structures that may be superimposed.

The model of this complex reanalysis is the structure of the simple fully finite clause, in which the initial auxiliary combines with the enclitic subject pronoun and is followed by the verbal complex (as, e.g., in the extrafocal clause of E28.b). The result of the change conforms exactly to that model.

Just in order to clarify a somewhat circuitous argument: The Yucatec data provide evidence for two interlinked grammaticalization paths:

<sup>18</sup> The suffix *-ah* is completive for transitive, incomplete for intransitive verbs.

<sup>19</sup> Sufficient evidence for this is provided, inter alia, by the documented history of the evolution of the progressive aspect in English and in substandard German; see Lehmann 1991, section 3.2.

(a) verb focus construction > focused progressive > (simple) progressive

(b) focused progressive of auxiliary 'go' > (simple) immediate future

Development (b) is not at stake here. It needed to be mentioned only because history is always more convoluted than diachronic typology: It so happened in the history of Yucatec Maya that the last stage of path (a) was barely reached (in Colonial constructions such as E31.c and d), and then the category of the progressive was renewed from a different source. The only trace left in the modern language of what was the Colonial focused and simple progressive is in the morphological structure of the immediate future (E32). The latter is, nevertheless, a useful example in our connection because it proves the complete loss of the underlying focus semantics.

The story of the Yucatec Maya focused progressive is a clear case of the grammaticalization of a construction that was motivated as a marked focus construction, whose information structure was leveled over time, but whose resulting structural properties still reflect the marked original information structure. It remains to add that the progressive aspect of other languages has a similar origin; cf., e.g., Güldemann 2003 for Bantu.

## 5 Conclusion

The relation between grammaticalization and information structure is twofold:

1. On the one hand, information structure is subject just to the same coding constraints as anything in language: Either it can be inferred from the linguistic or extralinguistic context, from the meaning of the sentence itself or on the basis of universal principles of semiosis; or else it must be coded. Coding is therefore necessary if a particular intended information structure differs from what would be the default for the given sentence or what would be inferred on the basis of universal principles or other clues.

As observed in section 3.1.1, the operations of information structure do not deal with concepts that may denote anything. The grammatical means employed in marking information structure are therefore immaterial (i.e. non-concrete) from the beginning. At a stage not too long past their genesis, such grammatical constructions as E7 do not involve a single (grammaticalized) lexical item. They make use of existent grammatical constructions such as presentational, identificational or anaphoric constructions and specialize these for the functions of information structure.

Such grammatical constructions that code information structure are subject to grammaticalization just like any other grammatical construction. As always in grammaticalization, the complexity level shrinks from the text level via the sentence and clause levels down to the phrase level. The lower the level of linguistic complexity, the more closely the coding unit corresponds to a single proposition. However, a single proposition cannot embody such a contrast as between presupposition vs. assertion, topic vs. comment, focus vs. presupposition; information structure requires at least two propositions that may be opposed in some way. As a consequence, the relief that is constitutive of the various dichotomies of information structure is leveled out under grammaticalization. What remains, in the end, is the simple topic-comment clause, where the topic is highly activated, i.e. represented by a clitic pronoun.

2. On the other hand, any construction that may serve as the input to some grammaticalization channel is shaped by information structure. Some lexical items in the construction will be focused, others will be topical or will at any rate be in the presupposition. If a certain construction is conventionally associated with a certain information structure, then such components of it that are topical or presupposed are particularly liable to become subordinate and reduced. So much can be generalized from the evolution of relative clauses analyzed in section 4.1. What may come rather unexpected is that the regular focus status of a component of a construction does not protect it from grammatical-

zation either; witness the pronominal interrogatives and the focused progressive analyzed in section 4.2f. The generalization here is that the particular information structure that shapes the initial construction is a presupposition for the course of grammaticalization that it takes; in other words, the same lexical-syntactic constellation in a different information structure would not embark on that grammaticalization path.

The complex interplay between grammaticalization and information structure thus consists in a mutual dependency: grammaticalization constrains and formalizes information structure, and information structure conditions and directs grammaticalization. The precise formulation of these conditions remains as a challenge for future work in grammaticalization.

## **Abbreviations**

### **Language**

AGr	Ancient Greek (Indo-European)
CYM	Colonial Yucatec Maya (Mayan)
Fr	French (Indo-European)
Ital	Italian (Indo-European)
Hit	Hittite (Indo-European)
Lat	Latin (Indo-European)
MYM	Modern Yucatec Maya (Mayan)
OFr	Old French (Indo-European)

**Interlinear glosses**

1, 2, 3	first, second, third person	INST	instrumental
ABL	ablative	INT	interrogative
ACC	accusative	INTROV	introversive (detransitive)
ADVR	adverbializer	LOC	locative
AOR	aoist	M	masculine
CMPL	completive	N	neuter
CONN	connective	NEG	negative
D3	distal deictic	NOM	nominative
DAT	dative	PL	plural
DEM	demonstrative	POSS	possessive
DEP	dependent	PRF	perfect
F	feminine	PRS	present
FUT	future	PST	past
GEN	genitive	PTCP	participle
IMP	imperative	REL	relative
INAN	inanimate	SBJ	subject
INCMPL	incompletive	SG	singular
IND	indefinite	SUBJ	subjunctive

**References**

- Arzápalo Marín, Ramón 1995, *Calepino de Motul. Diccionario Maya-Español. 3 tomos*. México D.F.: Universidad Nacional Autónoma de México.
- Banti, Giorgio 2005, "The typology of information structure: theoretical issues and open questions". Handout of a talk presented at University of Erfurt, 17.01.2005.
- Berretta, Monica 2002, "Ordini marcati dei costituenti maggiori di frase: una rassegna." Dal Negro, Silvia & Mortara Garavelli, Bice (eds.), *Temi e percorsi della linguistica. Scritti scelti di Monica Berretta*. Vercelli: Mercurio; 149-199.
- Coronel, Juan 1620, *Arte en lengua de maya, recopilado y enmendado por el P. Fray Joan Coronel*. México: D. Garrido.
- Givón, Talmy 1979, "From discourse to syntax: grammar as a processing strategy." Givón, Talmy (ed.), *Discourse and syntax*. New York etc.: Academic Press (Syntax and Semantics, 12); 81-112.
- Güldemann, Tom 2003, "Present progressive vis-à-vis predication focus in Bantu. A verbal category between semantics and pragmatics." *Studies in Language* 27:323-360.
- Hackstein, Olav 2004, "From discourse to syntax: the case of compound interrogatives in Indo-European and beyond." Jones-Bley, Karlene & Huld, Martin E. & Della Volpe, Angela & Dexter, Miriam Robbins (eds.), *Proceedings of the Fifteenth Annual UCLA Indo-European Conference. Los Angeles, November 7-8, 2003*. Washington: Institute for the Study of Man (Journal of Indo-European Monograph Series, 49); 257-298.
- Hyman, Larry M. 1984, "Form and substance in language universals." Butterworth, Brian & Comrie, Bernard & Dahl, Östen (eds.), *Explanations for language universals*. Berlin etc.: Mouton (*Linguistics* 21/1); 67-87.
- Lambrecht, Knud 1994, *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- LaPolla, Randy J. 1995, "Pragmatic relations and word order in Chinese." Downing, Pamela & Noonan, Michael (eds.), *Word order in discourse*. Amsterdam & Philadelphia: J. Benjamins (Typological Studies in Language, 30); 299-331.

- Lehmann, Christian 1984, *Der Relativsatz. Typologie seiner Strukturen - Theorie seiner Funktionen - Kompendium seiner Grammatik*. Tübingen: G. Narr (LUS, 2).
- Lehmann, Christian 1991, "Grammaticalization and related changes in contemporary German." Traugott, Elizabeth C. & Heine, Bernd (eds.), *Approaches to grammaticalization*. 2 vols. Amsterdam & Philadelphia: J. Benjamins (Typological Studies in Language, 19); II:493-535.
- Rouquier, Magali 2002, "Les interrogatives en 'qui/qu'est-ce qui/que' en ancien français et en moyen français." *Cahiers de grammaire* 27:97-120.
- Sankoff, Gillian & Brown, Penelope 1976, "The origins of syntax in discourse: A case study of Tok Pisin relatives." *Language* 52:631-666.
- Tai, James H.-Y. 1985, "Temporal sequence and Chinese word order." Haiman, John (ed.), *Iconicity in syntax. Proceedings of a symposium on iconicity in syntax, Stanford, June 24-6, 1983*. Amsterdam & Philadelphia: J. Benjamins (Typological Studies in Language, 6); 49-72.